# CausalSKyHop: Knowledge-Aware Causal Explanation of Dynamic GNNs via Higher-Order Semantic Reasoning

Jixuan Wu
Fujian Normal University
Fuzhou, China
w1986498047@163.com

Limei Lin*
Fujian Normal University
Fuzhou, China
linlimei@fjnu.edu.cn

Xiaoding Wang
Fujian Normal University
Fuzhou, China
wangdin1982@fjnu.edu.cn

Kunpeng Xu
McGill University
Montreal, Canada
kunpeng.xu@mail.mcgill.ca

Jie Wu
Temple University
Philadelphia, USA
jiewu@temple.edu

## Abstract

Dynamic Graph Neural Networks (DyGNNs) are widely used to model web-scale semantic-rich graph data (e.g., social networks, knowledge graphs), but their inability to explain predictions grounded in structured knowledge remains a challenge, especially when predictions rely on complex higher-order substructures. We propose CausalSKyHop, a semantic- and knowledge-aware framework that explains DyGNNs by uncovering causal higher-order patterns in evolving knowledge structures. To model the semantic fabric of the graph, CausalSKyHop incorporates a Higher-Order Structural Causal Model to capture multi-node knowledge dependencies, and uses contrastive learning to isolate semantically-meaningful causal relationships from spurious ones. A dynamic correlation module further separates persistent knowledge from evolving semantic contexts. Through knowledge-infused, structure-aware variational graph autoencoders, our method produces interpretable causal subgraphs that capture the dynamic flow of knowledge and semantics. Experimental evaluations on multiple web and knowledge-rich graph benchmarks demonstrate that CausalSKyHop consistently outperforms state-of-the-art explainable DyGNNs, achieving notable improvements in both explanation fidelity and downstream prediction accuracy. A detailed case study further illustrates how our method uncovers stable, semantically coherent causal pathways—in contrast to the fragmented explanations of baseline methods—providing intuitive evidence for its superior interpretability. This work establishes the critical role of explicit semantic and knowledge integration through higher-order causal reasoning for building transparent and trustworthy DyGNNs on the web.

## CCS Concepts

• **Theory of computation → Semantics and reasoning**; • **Computing methodologies → Causal reasoning and diagnostics**.

*Corresponding author: Limei Lin


*WWW '26, Dubai, United Arab Emirates.*

ACM ISBN 979-8-4007-2307-0/2026/04
https://doi.org/10.1145/3774904.3792242

## Keywords

Dynamic Graph Neural Networks; Causal Explanation; Higher-Order Semantic Reasoning; Knowledge-Aware Learning

**ACM Reference Format:**
Jixuan Wu, Limei Lin, Xiaoding Wang, Kunpeng Xu, and Jie Wu. 2026. CausalSKyHop: Knowledge-Aware Causal Explanation of Dynamic GNNs via Higher-Order Semantic Reasoning. In *Proceedings of the ACM Web Conference 2026 (WWW '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3774904.3792242

## 1 Introduction

The web is fundamentally a dynamic and interconnected tapestry of knowledge, where semantic-rich structures—from social networks [18] to knowledge graphs [19]—continuously evolve, carrying complex meaning and relationships. These temporal evolutions not only alter the topological structure but also transform the semantic contexts and knowledge dependencies that drive downstream tasks such as recommendation, anomaly detection in financial systems [1], and traffic flow prediction [26].

Dynamic Graph Neural Networks (DyGNNs) [25, 31, 32] have emerged as a powerful approach to capture both structural and temporal dependencies in evolving graphs. Recent models including WinGNN [34], DGA-GNN [4], and KEDGN [12] have achieved impressive predictive performance through sophisticated temporal embeddings and dynamic aggregation mechanisms. However, their black-box nature severely limits deployment in trust-sensitive applications, where explanations grounded in semantic and knowledge contexts are as crucial as predictions.

Existing explainability methods for GNNs—such as GNNExplainer [27], PGExplainer [11], OrphicX [9] and FORGE [17]—are primarily designed for static graphs and struggle to adapt to temporal settings. Despite the remarkable success of DyGNNs in modeling temporal graph data, their "black-box" nature severely hinders deployment in trust-sensitive applications. Existing explanation methods, such as DyGNNExplainer [33], DyExplainer [23], and DG-Explainer [10] are confined to first-order node-edge relationships, failing to capture the semantically rich higher-order structures that are ubiquitous in graphs. This limitation results in fragmented and semantically deficient explanations that are inadequate for uncovering the genuine causal mechanisms underlying model predictions.

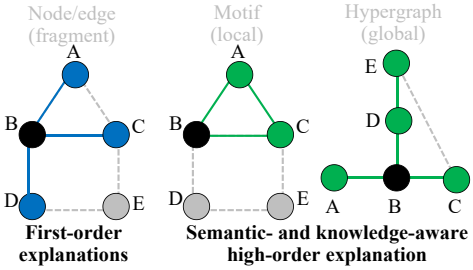



**Figure 1: From Fragmented First-Order to Coherent Higher-Order Explanations.**

To address the critical gap in higher-order semantic reasoning, we argue that effective explanations must capture both local semantic patterns and global complex dependencies inherent in dynamic knowledge structures (see Figure 1). This necessitates the integration of complementary higher-order structures: (1) **Motifs** capture recurring local subgraph patterns (e.g., triangles in social cliques) that serve as stable, interpretable causal units essential for semantic coherence. (2) **Hypergraphs** naturally model multi-node dependencies beyond pairwise interactions, preserving complex semantic relationships often lost in graph approximations and enabling knowledge-aware reasoning. The synergistic integration of these structures enables deeper higher-order semantic reasoning, overcoming the fragmentation of first-order methods.

To this end, we propose **CausalSKyHop**, a semantic- and knowledge-aware framework that explains DyGNNs by uncovering causal higher-order patterns in evolving knowledge structures. Our approach integrates explicit semantic reasoning with higher-order causal discovery to produce interpretable, semantically coherent explanations. The key contributions of this work are:

(1) **Causal Higher-Order Pattern Discovery.** We introduce a Higher-Order Structural Causal Model (HOSCM) to capture multi-node knowledge dependencies and leverage contrastive learning to isolate semantically meaningful causal relationships from spurious correlations. This enables the discovery of stable, interpretable causal pathways in dynamic graphs.

(2) **Semantic- and Knowledge-Aware Explanation Framework.** CausalSKyHop incorporates knowledge-infused, structure aware variational graph autoencoders to generate view-consistent explanations across diverse higher-order structures. Our theoretical analysis provides guarantees on the preservation of causal mechanisms during explanation generation.

(3) **Dynamic Correlation Disentanglement.** We design a dynamic correlation module with tailored disentanglement losses to separate persistent knowledge from evolving semantic contexts, enabling interpretable tracking of causal changes over time while filtering out spurious and static relationships.

(4) **Empirical Advancements.** Extensive evaluations on six web and knowledge-rich graph benchmarks demonstrate that CausalSKyHop consistently outperforms state-of-the-art explainable DyGNNs, improving explanation fidelity by 1.3%–7.1% and boosting downstream prediction accuracy by 4.6%–33.9% across all datasets while maintaining efficient scalability.

**Paper Organization.** The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 details the proposed **CausalSKyHop** framework. Section 4 presents extensive experimental results and analysis. Finally, Section 5 concludes the paper and discusses future directions.

## 2 Related Work

### 2.1 Interpretability of Graph Neural Networks

Existing GNN explanation methods can be broadly categorized into non-generative and generative approaches. Non-generative methods target instance-level explanations by identifying critical substructures via gradients [13], perturbations [15, 16], or mask optimization [27], with representative examples including Zorro [6], SubgraphX [29], and game-theoretic counterfactual methods based on semivalues (e.g., the Banzhaf value) that estimate feature contributions without additional training [3]. Generative methods, in contrast, learn distributions over explanatory subgraphs to improve automation and generalization, such as XGNN [28], PGM-Explainer [20], and GNNInterpreter [24].

Despite their prevalence, most methods remain structurally shallow and semantically unaware. They primarily model first-order node–edge relations without explicitly incorporating domain knowledge or semantic context, which is particularly limiting in knowledge-rich settings such as social networks and knowledge graphs where predictions often rely on higher-order semantic patterns.

DyGNNExplainer [33] is the first explainer tailored to dynamic GNNs. It disentangles spurious, static, and dynamic relations to account for temporal behaviors, yet it still overlooks higher-order structures and semantic dependencies that are fundamental to knowledge-intensive graphs. For semantic-rich web data, effective dynamic explanations should satisfy two criteria: fidelity [14], requiring the explanation subgraph to preserve the original model behavior, and semantic interpretability, requiring it to highlight meaningful knowledge patterns while filtering semantically irrelevant components.

### 2.2 Higher-Order Structures and Semantic Modeling

Higher-order structures—including motifs [30] and hypergraphs [5]—capture both local and global topological patterns, enhancing feature expressiveness beyond simple node-edge relationships. Recent works have leveraged such structures to improve graph learning: DyHGN [22] employs hypergraphs to enhance temporal graph modeling, while MOTIF-GNN [2] extracts higher-order patterns to boost predictive accuracy.

However, these approaches primarily focus on structural enhancement rather than semantic explicability. They lack mechanisms to ground higher-order patterns in domain knowledge or to establish causal relationships within semantic contexts. This represents a significant gap in knowledge-aware graph learning, particularly for dynamic environments where both structure and semantics evolve over time. Our work bridges this gap by introducing higher-order causal reasoning that explicitly incorporates structured knowledge to explain DyGNN predictions through semantically coherent pathways.

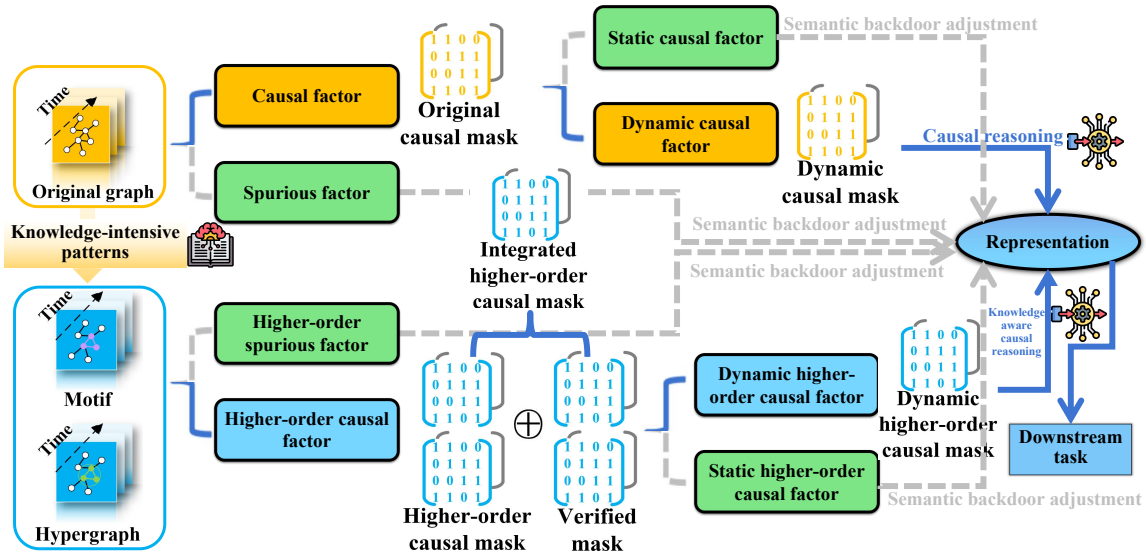



**Figure 2: The CausalSKyHop framework for explainable dynamic GNNs. The framework integrates (1) a Higher-Order Structural Causal Model (HOSCM) (see the flow directions of each block diagram in the figure) that captures knowledge-intensive patterns through causal reasoning, and (2) a temporal reasoning module that constructs dynamic masks while preserving knowledge coherence across time. Blue solid arrows indicate genuine causal paths, while gray dashed arrows represent backdoor paths that are blocked through semantic backdoor adjustment.**

## 3 Learning Knowledge-Aware Causal Higher-Order Patterns for Explainable Dynamic GNNs

This paper formulates the DyGNN model as $f = f_d \circ f_a$, where $f_a : G_{1:T} \to \mathbb{R}$ aggregates both structural topology and knowledge patterns from dynamic graphs $G_{1:T}$, producing high-dimensional representations, and $f_d : \mathbb{R} \to Y$ maps these representations to the label space $Y$. At each time step $t$, the input graph $G_t = (X_t, A_t)$ includes a node attribute matrix $X_t \in \mathbb{R}^{|V| \times D}$ (encoding node semantics and knowledge) and an adjacency matrix $A_t \in \mathbb{R}^{|V| \times |V|}$, with $V$ being the node set and $D$ the attribute dimension.

Interpreting knowledge-rich DyGNNs requires: *knowledge awareness* to identify meaningful causal patterns grounded in domain knowledge, *spatial interpretability* to locate critical graph components, and *temporal interpretability* to pinpoint evolution of knowledge contexts—all while remaining *model-agnostic* to explain any black-box model effectively. Our goal is to develop a generative interpreter that identifies knowledge-grounded causal subgraphs contributing to DyGNN predictions while maintaining fidelity. We focus on generating knowledge-aware spatiotemporal explanations—specifically, dynamic subgraph sets that capture evolving knowledge relationships—under a black-box setting where ground truth labels and model internals are unavailable.

### 3.1 Knowledge-Aware Causal Reasoning Framework

We propose **CausalSKyHop**, a causality-inspired multi-scale spatiotemporal explanation framework designed to perform knowledge-aware causal reasoning over dynamic graphs (see Figure 2). The core of our framework is the **Knowledge-Infused Higher-Order Structural Causal Model (HOSCM)**, which formally characterizes causal relationships among a set of key variables that jointly capture both the structural and semantic aspects of dynamic graphs. These variables include: the original graph $G$, the higher-order graph $HG$, spurious factors $S$, causal factors $C$, higher-order spurious factors $HS$, higher-order causal factors $HC$, dynamic factors $DC$, static factors $SC$, higher-order dynamic factors $DHC$, higher-order static factors $SHC$, node representation $R$, and the downstream task $Y$. A detailed exposition of the HOSCM is provided in Appendix A.

The causal links in the HOSCM represent both structural influences and semantic dependencies. Crucially, the higher-order factors—$HC$, $HS$, $DHC$, and $SHC$—are designed to explicitly capture knowledge-intensive patterns that extend beyond simple pairwise node-edge interactions. In our framework, "knowledge" refers to structured, higher-order patterns—specifically motifs and hypergraphs (see Appendix B for definitions)—that are explicitly extracted and leveraged. Meanwhile, "causal reasoning" denotes our methodology to systematically distinguish genuine causal relationships from spurious correlations, thereby yielding more faithful and interpretable explanations.

To operationalize this reasoning, we extract a variety of higher-order subgraph structures (motifs and hypergraphs) from each graph snapshot, and generate *causal soft masks* that encode knowledge aware patterns. These masks enable *semantic backdoor adjustment*, a mechanism that allows us to intervene on target causal and dynamic factors while controlling for confounding effects.

To ensure semantic coherence across higher-order structures, we employ a consistency loss combined with contrastive learning. A dynamic correlation module separates persistent and evolving

knowledge, while prediction and sparsity losses enhance explanation quality.

## 3.2 Semantic Backdoor Adjustment

Based on HOSCM, we identify four critical backdoor paths that introduce knowledge confounding into the learned representation:

(1) **Path 1**: $C \leftarrow G \rightarrow S \rightarrow R \rightarrow Y$ - spurious factor $S$ confounds the causal relationship between $C$ and $Y$.

(2) **Path 2**: $HC \leftarrow HG \rightarrow HS \rightarrow R \rightarrow Y$ - higher-order spurious factor $HS$ disrupts the knowledge connection between $HC$ and $Y$.

(3) **Path 3**: $DC \leftarrow C \rightarrow SC \rightarrow R \rightarrow Y$ - static component $SC$ interferes with the dynamic factor $DC$'s effect on $Y$.

(4) **Path 4**: $DHC \leftarrow HC \rightarrow SHC \rightarrow R \rightarrow Y$ - higher-order static factor $SHC$ introduces spurious dependency between $DHC$ and $Y$.

These backdoor paths contaminate the causal knowledge of learned representations, reducing robustness and interpretability. While do-calculus can theoretically eliminate confounding, practical application is challenging due to unobserved spurious and static factors. We propose a knowledge-infused masking strategy that explicitly suppresses non-causal components during training, effectively blocking these interference pathways while preserving knowledge coherence.

## 3.3 Disentangling Semantic Causal Relations

Given a dynamic graph sequence $G_{1:t} = (X_{1:t}, A_{1:t})$ for $1 \le t \le T$, we employ causal soft masks to decompose the graph into semantically meaningful components as follows.

(1) **Causal/Spurious Decomposition**: $M_t^C \in \mathbb{R}^{|V|\times|V|}$ partitions the graph into causal set $G_{1:t}^C = (X_{1:t}, A_{1:t} \odot M_{1:t}^C)$ and spurious set $G_{1:t}^S = (X_{1:t}, A_{1:t} \odot \overline{M}_{1:t}^C)$.

(2) **Dynamic/Static Decomposition**: $M_t^{DC}$ further decomposes causal set into dynamic causal set $G_{1:t}^{DC} = (X_{1:t}, A_{1:t} \odot M_{1:t}^C \odot M_{1:t}^{DC})$ and static causal set $G_{1:t}^{SC} = (X_{1:t}, A_{1:t} \odot M_{1:t}^C \odot \overline{M_{1:t}^{DC}})$.

We extend this mask-based decomposition to higher-order patterns that capture rich semantic relationships.

Let $\mathcal{S}_t = \{\text{motifs}, \text{hypergraphs}\}$ denote the set of higher-order structures, where each structure $s_t \in \mathcal{S}_t$ is assigned a causal mask $M_{s_t}^C$. These structure-specific masks aggregate into a unified higher-order causal mask $M_t^{HC}$, enabling extraction of higher-order causal set $G_{1:t}^{HC} = (X_{1:t}, A_{1:t} \odot M_{1:t}^{HC})$ and its complementary spurious set $G_{1:t}^{HS} = (X_{1:t}, A_{1:t} \odot \overline{M_{1:t}^{HC}})$. A higher-order dynamic mask $M_t^{DHC}$ further divides the higher-order causal subgraph into dynamic and static components.

Since these semantic sets are typically unobserved, our objective is to learn masks that effectively disentangle spurious, dynamic, and static relationships while preserving knowledge coherence.

*3.3.1 Knowledge-Enhanced Original Causal Mask Estimation.* We adopt a dynamic variational graph autoencoder enhanced with semantic constraints to estimate causal soft mask matrices $M_{1:T}^C$, capturing both temporal dependencies and structural semantics.

$$M_t^C = f_v(X_{1:t}, A_{1:t}; \vartheta_C), \tag{1}$$

where $f_v$ is the semantic-aware encoder with parameters $\vartheta_C$. The encoder-decoder architecture learns latent representations to infer dynamic subgraph structures with semantic fidelity (see Appendix C for details).

*3.3.2 Semantic-Rich Higher-Order Patterns Causal Mask Estimation.* We propose two knowledge-enhanced variational graph autoencoders (VGAEs) to estimate causal masks by capturing semantically meaningful higher-order interaction patterns: motifs and hypergraphs.

*Unified Semantic Framework.* For each structure $s_t \in \mathcal{S}_t$ with $s_t = \{v_1, v_2, \ldots, v_m\}$, we encode both structural and semantic information by concatenating node features and processing through a multilayer perceptron.

$$h_{s_t} = \text{MLP}_{\text{struct}}([x_{v_1}; x_{v_2}; \ldots; x_{v_m}]), \tag{2}$$

where $x_{v_i}$ captures the semantic attributes of node $v_i$. For each node $v_i$, we aggregate embeddings of all participating structures.

$$H_{\text{struct}_i} = \left(\sum\nolimits_{v_i \in s_t} h_{s_t}\right) / |\{s_t \in \mathcal{S}_t \mid v_i \in s_t\}|. \tag{3}$$

Node-level semantic features $H_{\text{node}_t} = \text{MLP}_{\text{node}}(X_t)$ are combined with structure-aware features.

$$H_{\text{combined}_t} = \text{MLP}_{\text{combined}}([H_{\text{node}_t}; H_{\text{struct}_t}]). \tag{4}$$

We then compute the latent distribution and decode to obtain the following causal mask.

$$z_t = \mu_t + \sigma_t \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad M_{\text{struct}_t}^C = \text{Decoder}(z_t). \tag{5}$$

*Semantic Specialization for Each Pattern.* (1) **Motifs**: Capture diverse **semantic interaction patterns** beyond simple triangles, defined as

$$\text{motifs} = \bigcup_{M_i \in \mathcal{M}} \{\{v_1, \ldots, v_m\} \mid G[\{v_1, \ldots, v_m\}] \cong M\}, \tag{6}$$

where each motif $M_i$ represents a meaningful semantic template.

(2) **Hypergraphs**: Model complex multi-way semantic relationships through hyperedges derived from motifs. The hypergraph Laplacian is defined as

$$L_{\text{hyper}} = I - D_v^{-1/2} H D_e^{-1} H^\top D_v^{-1/2} \tag{7}$$

This operator captures higher-order semantic dependencies unified with node representations.

The resulting causal masks $M_{\text{motif}_t}^C$ and $M_{\text{hyper}_t}^C$ encode structurally grounded semantic patterns.

*3.3.3 Multi-Level Semantic Mask Fusion.* We propose a multi-level structural mask fusion mechanism that integrates both topological information and higher-order semantic contexts through adaptive weights and enhancement factors.

*Semantic Mask Verification.* To address potential noise in individual higher-order masks, we introduce a cross-validation mechanism that ensures semantic consistency.

$$M_t^{\text{struct}_{\text{verified}}} = M_t^C \odot M_{\text{struct}_t}^C. \tag{8}$$

This step filters noisy components while preserving semantically meaningful patterns.

*Semantic Weight Computation.* We compute structure weights based on semantic importance rather than purely structural properties (detailed in Appendix D).

*Semantic-Aware Fusion.* Based on computed semantic weights, we perform weighted fusion as follows.

$$M_t^{HC} = w_1 \cdot M_t^{\text{motif}_{\text{verified}}} + w_2 \cdot M_t^{\text{hyper}_{\text{verified}}}. \tag{9}$$

We obtain causal adjacency matrices $A_{1:t}^C = A_{1:t} \odot M_{1:t}^C$ and higher-order causal adjacency matrices $A_{1:t}^{HC} = A_{1:t} \odot M_{1:t}^{HC}$, forming the basis for semantically coherent graph sets $G_{1:t}^C$ and $G_{1:t}^{HC}$.

Following the causal graph semantics, we compute dynamic masks using the same VGAE with specialized parameters.

$$M_t^{DC} = f_v(X_{1:t}, A_{1:t}^C \odot M_{1:t}^C; \vartheta_{DC}), M_t^{DHC} = f_v(X_{1:t}, A_{1:t}^C \odot M_{1:t}^{HC}; \vartheta_{DHC}). \tag{10}$$

This produces dynamic and static causal sets with corresponding higher-order variants, enabling comprehensive disentanglement of spurious, dynamic causal, and static causal relationships through semantically informed constraints.

## 3.4 Ensuring Semantic Consistency Across Structural Perspectives

To maintain semantic coherence across different higher-order structural masks, we propose the Higher-Order Structural Consistency Loss. This loss ensures that diverse structural perspectives capture complementary aspects of the underlying knowledge fabric rather than redundant information.

For time step $t$, with $\mathcal{S}_t = \{\text{motifs}, \text{hypergraphs}\}$ and the corresponding mask set $M_t$, the consistency loss is

$$\mathcal{L}_t^{\text{high}} = \mathcal{L}_t^{\text{verify}} + \beta \cdot \mathcal{L}_t^{\text{balance}}, \tag{11}$$

where $\mathcal{L}_t^{\text{verify}}$ ensures semantic verification consistency and $\mathcal{L}_t^{\text{balance}}$ maintains integrated mask balance (detailed in Appendix E). The overall higher-order consistency loss is:

$$\mathcal{L}_{\text{high}} = \frac{1}{T} \sum_{t=1}^{T} \mathcal{L}_t^{\text{high}}. \tag{12}$$

This loss serves as critical regularization that enforces **semantic alignment** across structural granularities, significantly improving robustness and interpretability.

*3.4.1 Semantic-Aware Decoupling of Spurious and Causal Relationships.* Since our targets are the causal and higher-order causal subgraph sets containing meaningful knowledge patterns, we treat spurious sets as semantic noise. We employ the pre-trained model's aggregation function $f_a(\cdot)$ to extract semantically rich embeddings.

$$\begin{aligned} &R_t = f_a(G_{1:t}), \quad R_t^C = f_a(G_{1:t}^C), \quad R_t^S = f_a(G_{1:t}^S), \\ &R_t^{HC} = f_a(G_{1:t}^{HC}), \quad R_t^{HS} = f_a(G_{1:t}^{HS}). \end{aligned} \tag{13}$$

We adopt contrastive learning to ensure semantic similarity between causal embeddings ($e_t^C$, $e_t^{HC}$) and raw embedding $e_t$, while maximizing distance from spurious embeddings ($e_t^S$, $e_t^{HS}$).

$$\mathcal{L}_c = \frac{1}{T} \sum_{t=1}^{T} \log \frac{\exp(s(\mathbf{e}_t, \mathbf{e}_t^C, \mathbf{e}_t^{HC})/\tau)}{\exp(s(\mathbf{e}_t, \mathbf{e}_t^C, \mathbf{e}_t^{HC})/\tau) + Z_t^C + Z_t^{HC}}, \tag{14}$$

where $Z_t^C$ and $Z_t^{HC}$ contrast causal against spurious semantics.

*3.4.2 Disentangling Semantic Dynamics from Static Knowledge.* We utilize Graph Convolutional Networks (GCNs) to extract dynamic and static relationships from corresponding causal subgraph sets (detailed in Appendix F). The key insight is that dynamic semantics evolve over time and can be inferred from historical contexts, while static knowledge remains context-independent.

$$\begin{aligned} &H_{1:(t-1)}^{DC} \to H_t^{DC}, H_{1:(t-1)}^{DHC} \to H_t^{DHC}, \\ &H_{1:(t-1)}^{SC} \perp H_t^{SC}, H_{1:(t-1)}^{SHC} \perp H_t^{SHC}. \end{aligned} \tag{15}$$

To ensure temporal semantic consistency, we introduce dynamic losses as follows.

$$\begin{aligned} \mathcal{L}_d &= \left( \sum_{t=2}^{T} d(f_a(G_{1:(t-1)}^{DC}), H_t^{DC}) \right) / (T-1), \\ \mathcal{L}_{hd} &= \left( \sum_{t=2}^{T} d(f_a(G_{1:(t-1)}^{DHC}), H_t^{DHC}) \right) / (T-1), \end{aligned} \tag{16}$$

where $d(\cdot, \cdot)$ measures distribution distance. These losses ensure that extracted dynamic relationships maintain semantic continuity over time, while static knowledge remains temporally invariant.

*3.4.3 Knowledge-Grounded Spatiotemporal Explanation.* From the structural causal model, both original and higher-order dynamic/static relationships jointly influence predictions through their semantic contributions. We treat dynamic relationships at time $t$ as interventions and define their causal effects as

$$\Delta H_t^{DC} = f_a(G_{1:t}^{DC}) - f_a(G_{1:(t-1)}^{DC}), \Delta H_t^{DHC} = f_a(G_{1:t}^{DHC}) - f_a(G_{1:(t-1)}^{DHC}). \tag{17}$$

We combine these dynamic causal effects with corresponding static relationships and employ a learnable aggregation mechanism to integrate temporally evolving semantics.

$$\begin{aligned} H_T &= \sum_{t=1}^{T} t_p(\Delta H_t^{DC} \odot H_t^{SC}) \Delta H_t^{DC} \odot H_t^{SC}, \\ H_T^H &= \sum_{t=1}^{T} t_p(\Delta H_t^{DHC} \odot H_t^{SHC}) \Delta H_t^{DHC} \odot H_t^{SHC}, \\ t_p(H) &= \text{Softmax}((\omega_\rho H)/\|\omega_\rho\|), \end{aligned} \tag{18}$$

where $\omega_\rho$ learns the temporal importance of semantic contributions, enhancing interpretability. Using a pre-trained classifier $f_d(\cdot)$, we predict labels with prediction losses $\mathcal{L}_p = l(f_d(H_T), y)$ and $\mathcal{L}_{hp} = l(f_d(H_T^H), y)$. To ensure human interpretability, we impose sparsity constraints on explained subgraphs.

$$\begin{aligned} \mathcal{L}_s &= \sum_{t=1}^{T} (\|A_t^C\|_1 + \|A_t^{DC}\|_1) / \|A_t^C\|_1, \\ \mathcal{L}_{hs} &= \sum_{t=1}^{T} (\|A_t^{HC}\|_1 + \|A_t^{DHC}\|_1) / \|A_t^{HC}\|_1. \end{aligned} \tag{19}$$

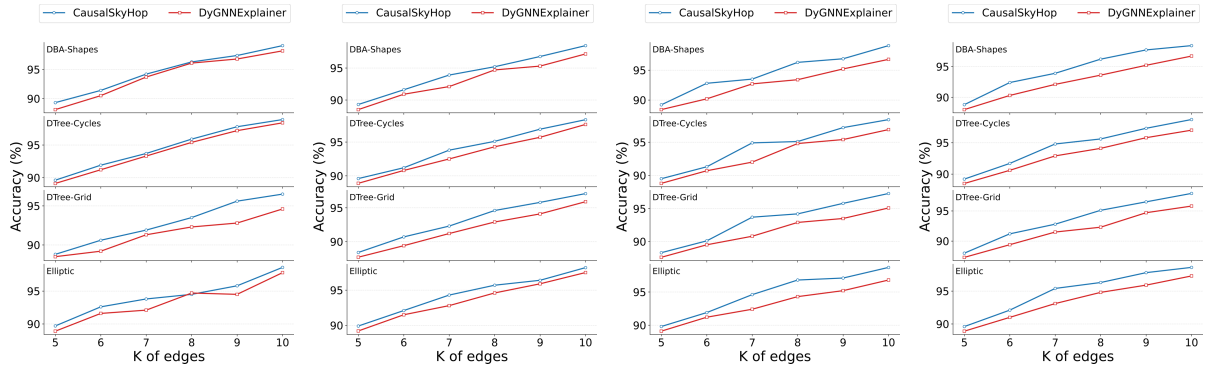
We jointly learn optimal interpretable causal subgraphs through unified optimization.

$$\begin{aligned} \min_{(\vartheta,\omega)} \mathcal{L}(\vartheta, \omega) = &\xi_1 \mathcal{L}_c + \xi_2 \mathcal{L}_s + \xi_3 \mathcal{L}_{hs} + \xi_4 \mathcal{L}_p + \xi_5 \mathcal{L}_{hp} \\ &+ \xi_6 \mathcal{L}_d + \xi_7 \mathcal{L}_{hd} + \xi_8 \mathcal{L}_{high}, \end{aligned} \tag{20}$$

where $\vartheta$ and $\omega$ encompass all model parameters, and $\xi_1$-$\xi_8$ balance different semantic objectives.

# 4 Experiments

Detailed descriptions of the datasets, baseline methods, and implementation settings are provided in Appendix G.




(a) t=2 (b) t=4 (c) t=6 (d) t=8

**Figure 3: Interpretability analysis of CausalSKyHop versus baselines. Our method produces more sparse and semantically coherent explanations by leveraging higher-order structures (motifs and hypergraphs), effectively isolating causal pathways from spurious correlations.**

## 4.1 Explanation Fidelity and Interpretability Evaluations

**Table 1: The number of high-order structures that appear in different datasets and at different time steps.**

| Number | $t$ | DBA-S | DTree-C | DTree-G | Ell | Meme |
|---|---|---|---|---|---|---|
| No. of motifs | 1 | 9240 | 2115 | 5317 | 2529 | 118 |
| | 2 | 9240 | 2114 | 5330 | 1604 | 138 |
| | 3 | 9240 | 2115 | 5329 | 3012 | 95 |
| | 4 | 9240 | 2112 | 5335 | 2491 | 95 |
| | 5 | 9239 | 2115 | 5322 | 2921 | 68 |
| No. of hypergraphs | 1 | 5 | 5 | 5 | 163 | 45 |
| | 2 | 5 | 5 | 5 | 208 | 52 |
| | 3 | 5 | 5 | 5 | 485 | 34 |
| | 4 | 5 | 5 | 5 | 345 | 31 |
| | 5 | 5 | 5 | 5 | 517 | 25 |

*4.1.1 Explanation Interpretability.* Figure 3 provides a comparative analysis of explanation quality, where **CausalSKyHop** generates significantly more interpretable and sparse causal subgraphs than competing methods. Our framework's integration of higher-order structural reasoning enables it to isolate semantically meaningful causal pathways while filtering out spurious correlations. The use of motifs (m) captures local stable patterns, while hypergraphs (h) model global multi-node dependencies—together providing a comprehensive semantic fabric for explanation. This structural synergy allows our method to maintain high fidelity to the original model's behavior while producing human-intelligible explanations that are both concise and semantically coherent, addressing a key limitation of fragmented explanations produced by first-order methods [7].

*4.1.2 Explanation Fidelity.* As quantitatively demonstrated in Figure 4, **CausalSKyHop** achieves state-of-the-art performance across all benchmark datasets, surpassing existing explainable DyGNNs by a significant margin. This consistent superiority stems from our framework's ability to systematically uncover causal higher-order patterns in evolving knowledge structures. Unlike baselines that operate on first-order neighborhoods, our method leverages motifs and hypergraphs to capture semantically rich substructures that are critical for accurate predictions. The performance gains—ranging from 1.3% to 7.1% in explanation fidelity—provide strong empirical evidence that explicit modeling of higher-order semantic dependencies is essential for both interpretability and predictive performance in dynamic graph learning.

*4.1.3 Persistence of High-order Structures.* The persistent presence of higher-order structures over time (Table 1) indicates that incorporating both motifs and hypergraphs into the framework is necessary. In datasets such as DBA-S, DTree-C, and DTree-G, these structures remain stable and abundant at every time step, suggesting that they are not transient phenomena but long-term organizational principles of the graphs. Even in more dynamic datasets such as Ell and Meme, they consistently appear and never disappear. Therefore, explicitly modeling these persistent higher-order structures helps CausalSkyHop capture the core "skeleton" of the graph, enabling more stable and robust explanations under temporal fluctuations.

## 4.2 Comprehensive Performance Evaluation

*4.2.1 Performance Comparison with Baselines.* As shown in Figure 5 (a), CausalSkyHop achieves remarkable performance gains across both benchmark datasets. On the DBA-Shapes dataset, our method reaches 47.2% accuracy, outperforming the next best baseline (Target DyGNNEx at 44.6%) by 2.6 percentage points. The improvement is even more pronounced on the Elliptic dataset, where CausalSkyHop achieves 95.8% accuracy compared to 89.2% for Target DyGNNEx, representing a 6.6 percentage point improvement.

This consistent superiority across diverse graph types–from synthetic structural patterns in DBA-Shapes to real-world financial transactions in Elliptic–validates the effectiveness of our higher-order causal reasoning approach. The significant performance gap, particularly on the complex Elliptic dataset, underscores CausalSkyHop's capability to capture semantically meaningful patterns in dynamic graph structures.

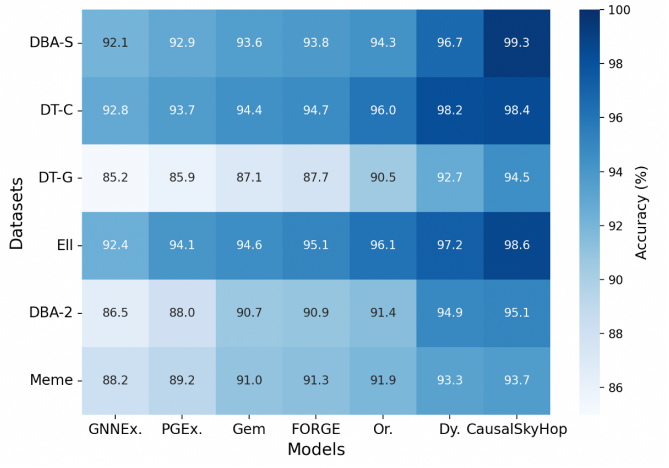
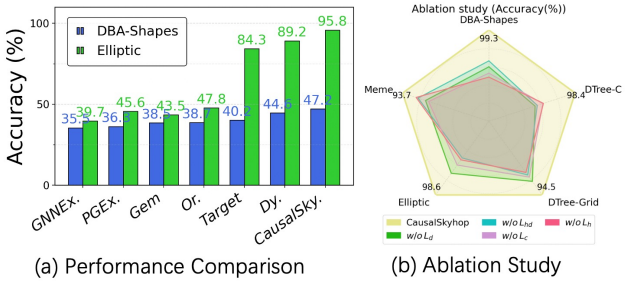



**Figure 4: Comparative performance analysis of explainable DyGNN methods across multiple benchmarks. The heatmap visualization highlights the consistent superiority of CausalSkyHop in terms of explanation fidelity across all evaluated datasets. The color intensity corresponds to performance metrics, with darker shades indicating higher scores.**




(a) Performance Comparison (b) Ablation Study

**Figure 5: (a) Performance comparison with state-of-the-art methods on two representative benchmarks, demonstrating superior overall accuracy. (b) Ablation study across five datasets, revealing the critical contribution of each component. The color intensity in the heatmap corresponds to accuracy levels.**

*4.2.2 Ablation Study on Component Contributions.* Figure 5(b) presents ablation results across multiple datasets, evaluating the contribution of each core component: (1) **Contrastive Loss ($\mathcal{L}_c$)**: Removing $\mathcal{L}_c$ causes the most significant performance degradation across all datasets, confirming its crucial role in distinguishing causal relationships from spurious correlations. (2) **Dynamic Loss ($\mathcal{L}_d$)**: Ablating $\mathcal{L}_d$ leads to noticeable performance drops, particularly in temporally dynamic datasets, validating the importance of separating persistent and evolving information. (3) **Higher-order Dynamic Loss ($\mathcal{L}_{hd}$)**: Removing $\mathcal{L}_{hd}$ significantly degrades performance due to the inability to capture the temporal evolution of higher-order semantics, indicating that jointly modeling their temporal consistency and dynamics is necessary. (4) **Higher-order Consistency Loss ($\mathcal{L}_h$)**: While having relatively smaller impact,

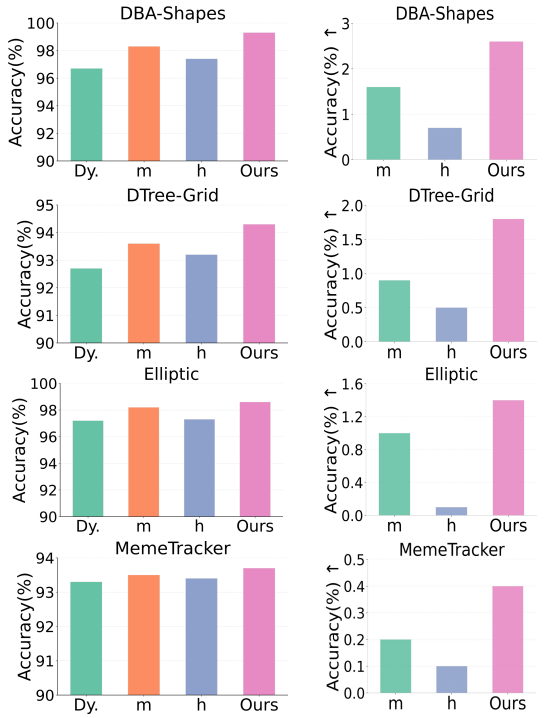



**Figure 6: Ablation study on the contribution of different higher-order structures (motif: m, hypergraph: h) and loss components.**

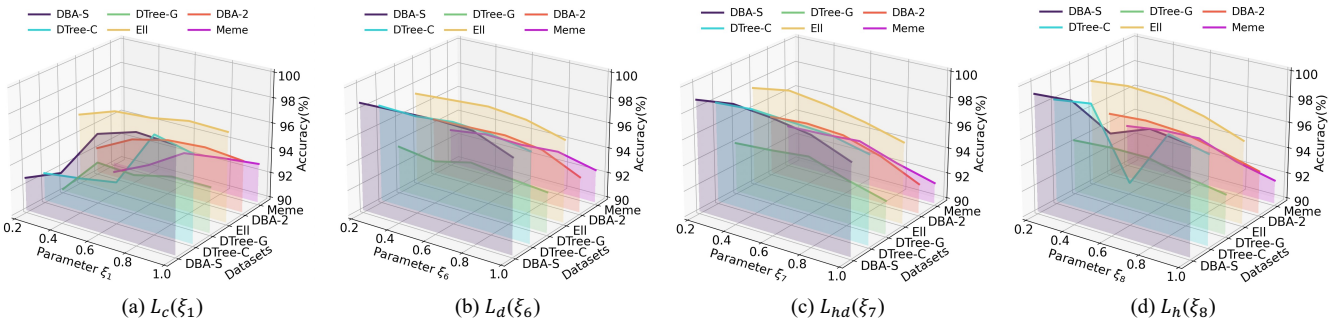
removing $\mathcal{L}_h$ still causes performance deterioration, emphasizing the value of coordinated multi-structural integration.

The full model consistently outperforms all ablated variants, with the performance ranking (Full > w/o $\mathcal{L}_{hd}$ > w/o $\mathcal{L}_d$ > w/o $\mathcal{L}_h$ > w/o $\mathcal{L}_c$) remaining stable across datasets. This demonstrates that all components work synergistically for optimal performance.

The significant drop when removing $\mathcal{L}_c$ suggests that existing methods may be confounded by spurious correlations, while our approach successfully isolates genuine causal mechanisms in dynamic graphs.

## 4.3 Ablation Study

Figure 6 reports an ablation study on higher-order structures in CausalSKyHop, where we evaluate motifs (m), hypergraphs (h), and their combination. The results show that introducing either structure alone improves over the baseline without higher-order structures, and motifs yield more pronounced gains because they better capture stable local patterns. Combining both structures achieves the best performance, indicating that multi-structural fusion provides a more comprehensive causal modeling capability. Overall, motifs focus on local recurrent patterns, while hypergraphs model global multi-node dependencies. Their synergy offers a more complete semantic support for causal explanations, alleviates the fragmentation of first-order methods, and delivers robust interpretability across diverse graph structures.




(a) $L_c(\xi_1)$ (b) $L_d(\xi_6)$ (c) $L_{hd}(\xi_7)$ (d) $L_h(\xi_8)$

**Figure 7: Parameter sensitivity analysis of different coefficients. Each subfigure varies one parameter; the model remains robust and peaks near the chosen values.**

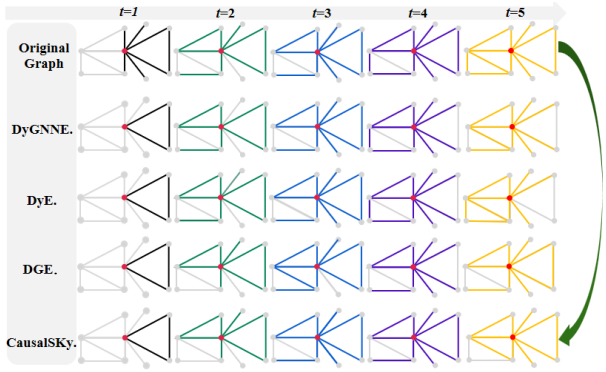



**Figure 8: Case study on DBA-Shapes, visualizing the causal explanations generated by CausalSKyHop and baselines. Our method consistently identifies the complete "fish" motif structure across time steps, demonstrating its ability to provide stable and semantically coherent causal pathways.**

## 4.4 Sensitivity Study

To evaluate the influence of key hyperparameters on model performance, we conducted parameter sensitivity experiments on the coefficients $\xi_1$, $\xi_6$, $\xi_7$, and $\xi_8$ in Eq. (18). Each parameter was independently varied within $0.2, 0.4, 0.6, 0.8, 1.0$ while keeping the remaining coefficients fixed at 0.2. Specifically, we first tuned $\xi_1$ under fixed settings of $\xi_6 = \xi_7 = \xi_8 = 0.2$ and selected its optimal value according to the validation performance. Then, $\xi_1$ was fixed to its optimal value while tuning $\xi_6$, and this process was repeated sequentially for $\xi_7$ and $\xi_8$. The corresponding results are shown in Figure 7, where each subfigure depicts the effect of varying a single parameter. As illustrated in the figure, the model maintains stable and consistent performance across a wide range of parameter values and achieves optimal performance near the selected values, demonstrating the robustness of the proposed framework.

## 4.5 Case Study

The temporal case study in Figure 8 offers qualitative evidence of **CausalSKyHop**'s ability to provide stable and semantically coherent explanations across time. In the DBA-Shapes dataset, our method consistently identifies the complete "fish" motif—a complex higher-order structure—across all four time steps, accurately reconstructing its semantic layout around the target node. In contrast, DyGNNExplainer produces incomplete and inconsistent substructures, often including irrelevant edges. This visual comparison demonstrates how our higher-order causal modeling captures structurally stable and temporally persistent patterns that are essential for trustworthy explanations. The case study thus validates that **CausalSKyHop** not only identifies causally relevant substructures but also maintains their semantic integrity over time, a critical requirement for explaining dynamic graph models in real-world applications. The entire CausalSKyHop analysis process is shown in Appendix H.

## 5 Conclusion

We proposed **CausalSKyHop**, a knowledge-aware framework that explains Dynamic Graph Neural Networks (DyGNNs) through higher-order causal reasoning. Our approach enhances DyGNN interpretability by explicitly modeling semantic dependencies in dynamic graphs, with three key contributions: (1) a Higher-Order Structural Causal Model capturing multi-node patterns; (2) semantic-aware masking for identifying meaningful causal relationships; and (3) temporal reasoning separating persistent knowledge from evolving contexts. Experimental results demonstrate consistent improvements over state-of-the-art methods in both explanation fidelity and prediction accuracy. Qualitatively, our method generates semantically coherent causal pathways, offering more intuitive explanations than fragmented baseline outputs. This work highlights the crucial role of semantic and knowledge integration in building transparent DyGNNs. Future work will explore adaptive structure selection, extension to heterogeneous graphs, and scaling to web-scale dynamic graphs.

## Acknowledgments

This work is supported by the Natural Science Foundation of Fujian Province (No. 2024J09032, No. 2025J01379, No. 2025H0043, No. 2025J02019), the Joint Funds for the Innovation of Science and Technology of Fujian Province (No. 2024Y9491).

## References


[1] Erik R. Altman, Jovan Blanuša, Luc von Niederhäusern, Beni Egressy, Andreea Anghel, and Kubilay Atasu. 2023. Realistic Synthetic Financial Transactions for Anti-Money Laundering Models. In *Advances in Neural Information Processing Systems*, Vol. 36. Curran Associates, Inc., New Orleans, LA, USA. https://proceedings.neurips.cc/paper_files/paper/2023/hash/5f38404edff6f3f642d6fa5892479c42-Abstract-Datasets_and_Benchmarks.html

[2] Xuexin Chen, Ruichu Cai, Yuan Fang, Min Wu, Zijian Li, and Zhifeng Hao. 2024. Motif Graph Neural Network. *IEEE Transactions on Neural Networks and Learning Systems* 35, 10 (2024), 14833–14847. https://doi.org/10.1109/TNNLS.2023.3281716

[3] Chirag Chhablani, Sarthak Jain, Akshay Channesh, Ian A. Kash, and Sourav Medya. 2024. Game-theoretic Counterfactual Explanation for Graph Neural Networks. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. Association for Computing Machinery, Singapore, 503–514. https://doi.org/10.1145/3589334.3645419

[4] Mingjiang Duan, Tongya Zheng, Yang Gao, Gang Wang, Zunlei Feng, and Xinyu Wang. 2024. DGA-GNN: Dynamic Grouping Aggregation GNN for Fraud Detection. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI '24)*. AAAI Press, Vancouver, BC, Canada, 11820–11828. https://doi.org/10.1609/aaai.v38i10.29067

[5] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. 2019. Hypergraph Neural Networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI '19)*. AAAI Press, Honolulu, HI, USA, 3558–3565. https://doi.org/10.1609/aaai.v33i01.33013558

[6] Thorben Funke, Megha Khosla, Mandeep Rathee, and Avishek Anand. 2023. Zorro: Valid, Sparse, and Stable Explanations in Graph Neural Networks. *IEEE Transactions on Knowledge and Data Engineering* 35, 8 (2023), 8687–8698. https://doi.org/10.1109/TKDE.2022.3201170

[7] Antoine Ledent and Peng Liu. 2025. Explainable Neural Networks with Guarantee: A Sparse Estimation Approach. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI '25)*. AAAI Press, Philadelphia, PA, USA, 18044–18052. https://doi.org/10.1609/aaai.v39i17.33985

[8] Wanyu Lin, Hao Lan, and Baochun Li. 2021. Generative Causal Explanations for Graph Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*. PMLR, 6666–6679. https://proceedings.mlr.press/v139/lin21d.html

[9] Wanyu Lin, Hao Lan, Hao Wang, and Baochun Li. 2022. OrphicX: A Causality-Inspired Latent Variable Model for Interpreting Graph Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '22)*. IEEE, New Orleans, LA, USA, 13719–13728. https://doi.org/10.1109/CVPR52688.2022.01336

[10] Yezi Liu, Jiaxuan Xie, and Yanning Shen. 2025. DGExplainer: Explaining Dynamic Graph Neural Networks via Relevance Back-propagation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2025)*. International Joint Conferences on Artificial Intelligence Organization, Montreal, Canada, 3135–3143. https://doi.org/10.24963/ijcai.2025/349

[11] Dongsheng Luo, Wei Cheng, Dongkuan Xu, Wenchao Yu, Bo Zong, Haifeng Chen, and Xiang Zhang. 2020. Parameterized Explainer for Graph Neural Network. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates, Inc., Virtual, 19620–19631. https://proceedings.neurips.cc/paper/2020/hash/e37b08dd3015330dcbb5d6663667b8b8-Abstract.html

[12] Yicheng Luo, Zhen Liu, Linghao Wang, Binquan Wu, Junhao Zheng, and Qianli Ma. 2024. Knowledge-Empowered Dynamic Graph Network for Irregularly Sampled Medical Time Series. In *Advances in Neural Information Processing Systems*, Vol. 37. Curran Associates, Inc., Vancouver, BC, Canada, 67172–67199. https://doi.org/10.52202/079017-2144

[13] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. 2019. Explainability Methods for Graph Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '19)*. IEEE, Long Beach, CA, USA, 10772–10781. https://doi.org/10.1109/CVPR.2019.01103

[14] Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why Should I Trust You?: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, San Francisco, CA, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[15] Michael Sejr Schlichtkrull, Nicola De Cao, and Ivan Titov. 2021. Interpreting Graph Neural Networks for NLP With Differentiable Edge Masking. In *Proceedings of the Ninth International Conference on Learning Representations (ICLR 2021)*. OpenReview.net, Virtual. https://openreview.net/forum?id=WznmQa42ZAx

[16] Thomas Schnake, Oliver Eberle, Jonas Lederer, Shinichi Nakajima, Kristof T. Schütt, Klaus-Robert Müller, and Grégoire Montavon. 2022. Higher-Order Explanations of Graph Neural Networks via Relevant Walks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 11 (2022), 7581–7596. https://doi.org/10.1109/TPAMI.2021.3115452

[17] Akshit Sinha, Sreeram Vennam, Charu Sharma, and Ponnurangam Kumaraguru. 2025. Higher Order Structures for Graph Explanations. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI '25)*. AAAI Press, Philadelphia, PA, USA, 20514–20521. https://doi.org/10.1609/aaai.v39i19.34260

[18] Youchen Sun, Zhu Sun, Yingpeng Du, Jie Zhang, and Yew Soon Ong. 2025. Model-Agnostic Social Network Refinement with Diffusion Models for Robust Social Recommendation. In *Proceedings of the ACM Web Conference 2025 (WWW '25)*. Association for Computing Machinery, Sydney, NSW, Australia, 370–378. https://doi.org/10.1145/3696410.3714683

[19] Lionel Tailhardat, Benjamin Stach, Yoan Chabot, and Raphaël Troncy. 2024. Graphameleon: Relational Learning and Anomaly Detection on Web Navigation Traces Captured as Knowledge Graphs. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24)*. Association for Computing Machinery, Singapore, 1103–1106. https://doi.org/10.1145/3589335.3651447

[20] Minh N. Vu and My T. Thai. 2020. PGM-Explainer: Probabilistic Graphical Model Explanations for Graph Neural Networks. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates, Inc., Virtual, 12225–12235. https://proceedings.neurips.cc/paper/2020/hash/8fb134f258b1f7865a6ab2d935a897c9-Abstract.html

[21] Junxiang Wang, Junji Jiang, and Liang Zhao. 2022. An Invertible Graph Diffusion Neural Network for Source Localization. In *Proceedings of the ACM Web Conference 2022 (WWW '22)*. Association for Computing Machinery, Lyon, France, 1058–1069. https://doi.org/10.1145/3485447.3512155

[22] Tianci Wang, Yantong Lai, Gaode Chen, Ruohan Wang, Jiahui Shen, and Ji Xiang. 2023. A Dynamic-Aware Heterogeneous Graph Neural Network for Next POI Recommendation. In *Proceedings of the 20th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2023), Part I*. Springer, Jakarta, Indonesia, 313–326. https://doi.org/10.1007/978-981-99-7019-3_30

[23] Tianchun Wang, Dongsheng Luo, Wei Cheng, Haifeng Chen, and Xiang Zhang. 2025. DyExplainer: Self-explainable Dynamic Graph Neural Network with Sparse Attentions. *ACM Transactions on Knowledge Discovery from Data* 19, 4 (2025), 1–21. https://doi.org/10.1145/3729173

[24] Xiaoqi Wang and Han-Wei Shen. 2023. GNNInterpreter: A Probabilistic Generative Model-Level Explanation for Graph Neural Networks. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR 2023)*. OpenReview.net, Kigali, Rwanda. https://openreview.net/forum?id=rqq6Dh8t4d

[25] Xi Wang, Qianzhen Zhang, Deke Guo, and Xiang Zhao. 2023. A Survey of Continuous Subgraph Matching for Dynamic Graphs. *Knowledge and Information Systems* 65, 3 (2023), 945–989. https://doi.org/10.1007/s10115-022-01753-x

[26] Yan Xu, Yu Lu, Changtao Ji, and Qiyuan Zhang. 2023. Adaptive Graph Fusion Convolutional Recurrent Network for Traffic Forecasting. *IEEE Internet of Things Journal* 10, 13 (2023), 11465–11475. https://doi.org/10.1109/JIOT.2023.3244182

[27] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. GNNExplainer: Generating Explanations for Graph Neural Networks. In *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates, Inc., Vancouver, BC, Canada, 9240–9251. doi:10.52202/079017-2144

[28] Hao Yuan, Jiliang Tang, Xia Hu, and Shuiwang Ji. 2020. XGNN: Towards Model-Level Explanations of Graph Neural Networks. In *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '20)*. Association for Computing Machinery, San Diego, CA, USA, 430–438. https://doi.org/10.1145/3394486.3403085

[29] Hao Yuan, Haiyang Yu, Jie Wang, Kang Li, and Shuiwang Ji. 2021. On Explainability of Graph Neural Networks via Subgraph Explorations. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*. PMLR, 12241–12252. https://proceedings.mlr.press/v139/yuan21c.html

[30] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. 2021. Motif-based Graph Self-Supervised Learning for Molecular Property Prediction. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., Virtual, 15870–15882. https://proceedings.neurips.cc/paper/2021/hash/85267d349a5e647ff0a9edcb5ffd1e02-Abstract.html

[31] Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Zhou Qin, and Wenwu Zhu. 2022. Dynamic Graph Neural Networks Under Spatio-Temporal Distribution Shift. In *Proceedings of the Thirty-Sixth Annual Conference on Neural Information Processing Systems (NeurIPS 2022)*. Curran Associates, Inc., New Orleans, LA, USA, 6074–6089. http://papers.nips.cc/paper_files/paper/2022/hash/2857242c9e97de339ce642e75b15ff24-Abstract-Conference.html

[32] Zeyang Zhang, Xin Wang, Ziwei Zhang, Zhou Qin, Weigao Wen, Hui Xue, Haoyang Li, and Wenwu Zhu. 2023. Spectral Invariant Learning for Dynamic Graphs under Distribution Shifts. In *Advances in Neural Information Processing Systems*, Vol. 36. Curran Associates, Inc., New Orleans, LA, USA, 6619–6633. http://papers.nips.cc/paper_files/paper/2023/hash/154b90fcc9ba3dee96779c05c3108908-Abstract-Conference.html

[33] Kesen Zhao and Liang Zhang. 2024. Causality-Inspired Spatial-Temporal Explanations for Dynamic Graph Neural Networks. In *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*. OpenReview.net, Vienna, Austria. https://openreview.net/forum?id=AJBkfwXh3u

[34] Yifan Zhu, Fangpeng Cong, Dan Zhang, Wenwen Gong, Qika Lin, Wenzheng Feng, Yuxiao Dong, and Jie Tang. 2023. WinGNN: Dynamic Graph Neural Networks with Random Gradient Aggregation Window. In *Proceedings of the Twenty-Ninth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*.

Association for Computing Machinery, Long Beach, CA, USA, 3650–3662. https://doi.org/10.1145/3580305.3599551

# A Explanation of HOSCM

**Key Explanations Regarding HOSCM**. We adopt a Higher-order structural Causal Model (HOSCM) to formally describe the causal mechanisms in dynamic graph neural networks. The notations and relationships are defined as follows:

- $S \leftarrow G \rightarrow C$: The original graph $G$ gives rise to both the causal factors $C$ and the spurious factors $S$, the latter typically resulting from spurious correlations or inherent biases in the data.
- $HS \leftarrow HG \rightarrow HC$: The higher-order dynamic graph $HG$ gives rise to both the higher-order causal factors $HC$ and the higher-order spurious factors $HS$, the latter typically resulting from spurious correlations or inherent biases in the data.
- $S \rightarrow R \leftarrow C$: The high-dimensional representation $R$ is jointly influenced by spurious factors $S$ and causal factors $C$ extracted from the original graph. The spurious factors $S$ typically arise from correlation biases or structural noise, while the causal factors $C$ capture task-relevant causal information essential for accurate reasoning.
- $HS \rightarrow R \leftarrow HC$: In the context of higher-order structures, the representation $R$ is further affected by higher-order spurious factors $HS$ and higher-order causal factors $HC$. The spurious components $HS$ often originate from pseudo-correlations present in complex structural patterns such as motifs or cliques, whereas the higher-order causal factors $HC$ provide complementary semantic cues beyond what standard causal representations can offer.
- $DC \rightarrow R \leftarrow SC$: In the structure of dynamic graphs, the causal component $C$ is influenced by a combination of time-varying dynamics $DC$ and temporally invariant static factors $SC$.
- $DHC \rightarrow R \leftarrow SHC$: In the structure of dynamic graphs, the higher-order causal component $HC$ is influenced by a combination of higher-order dynamic factors $DHC$ and higher-order static factors $SHC$.
- $R \rightarrow Y$: The learned representation $R$ serves as the basis for downstream predictive tasks, including inference at the node or whole-graph level.

# B Definition of Higher-Order Patterns

**(Motif)** A motif is a small-scale, recurring structural pattern in a graph, used to characterize common local connectivity structures. Let $\mathcal{M} = \{M_i\}$ be a set of motif templates, where each template $M = (V_M, E_M)$ represents a meaningful semantic pattern. Given a (snapshot) graph $G = (V, E)$, a motif instance is a node set $\{v_1, \ldots, v_m\}$ such that the induced subgraph $G[\{v_1, \ldots, v_m\}]$ is isomorphic to a template $M \in \mathcal{M}$. The formal set definition of motif instances is given in Eq. (6) in the main text.

**(Hypergraph)** A hypergraph is a generalization of a standard graph in which an edge, called a *hyperedge*, can connect more than two nodes. Formally, a hypergraph is defined as $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a set of nodes and $\mathcal{E} \subseteq 2^{\mathcal{V}}$ is a set of hyperedges, each of which is a non-empty subset of $\mathcal{V}$.

# C Implementation Details of Original Causal Mask Estimation

The encoder utilizes the posterior probability to embed nodes into low-dimensional latent vector representations, which can be expressed as

$$\begin{aligned} q(H_t|G_{1:t}) &= \prod_{i=1}^{N} q(h_{t,i}|G_{1:t}), \\ q(h_{t,i}|G_{1:t}) &= \mathcal{N}(h_{t,i}|\mu_{t,i}, \mathrm{diag}(\sigma_{t,i}^2)), \end{aligned} \tag{21}$$

where $H_t$ represents the latent matrix, $\mu_t$ and $\sigma_t$ are means and variances of node latent embeddings learned by GCN$\mu(G_t)$ and GCN$\sigma(G_t)$ with different parameters, $h_{t,i}$, $\mu_{t,i}$, and $\sigma_{t,i}$ are the $i$-th columns of $H_t$, $\mu_t$, and $\sigma_t$, respectively. We employ the reparameterization trick to enable gradient-based optimization of the sampling process. The decoder then uses these latent representations to produce dynamically interpretable subgraphs as follows.

$$\begin{aligned} p(M_t^C|H_t) &= \prod_{i=1}^{N} \prod_{j=1}^{N} p(M_{t,ij}^C|h_{t,i}, h_{t,j}), \\ p(M_{t,ij}^C = 1|h_{t,i}, h_{t,j}) &= g(h_{t,i}, h_{t,j}), \end{aligned} \tag{22}$$

where $M_{t,ij}^C$ represents the $i$-th row and the $j$-th column element of $M_t^C$, indicating the probability that the edge $(v_t^i, v_t^j)$ exists in the causal graph set at time $t$, and $g(\cdot, \cdot)$ represents the probability.

# D Structure Weight Computation

To effectively fuse the verified higher-order masks, we propose a weight computation method based on structural importance. The structure weight vector is defined as $\mathbf{w} = [w_1, w_2]$, where $w_1, w_2$ represent the individual weights of the two higher-order structures. The detailed computational steps for determining the weight vector are as follows:

*Constructing the Correlation Matrix.* We first construct a correlation matrix $S \in \mathbb{R}^{2\times 2}$, where each entry $S_{ij}$ measures the similarity between the $i$-th and $j$-th types of higher-order structures:

$$S_{ij} = \mathrm{sim}(M_i^v, M_j^v), \quad (i \neq j), \tag{23}$$

where $M_i^v$ and $M_j^v$ are the verified higher-order masks corresponding to structures $i$ and $j$, respectively.

*Reducing Redundant Structure Weights.* To address redundancy, we compute the redundancy score of each structure by averaging its similarity with the other structures:

$$p_i = \frac{1}{K-1} \sum_{j \neq i} S_{ij}, \quad K = 2, \tag{24}$$

where $\mathbf{p} = [p_1, p_2]$ represents the average redundancy of each structure.

*Adjusting the Weights.* Using the redundancy penalty coefficient $\alpha$, we adjust the structure weights $\mathbf{w}'$ by reducing the weights of redundant structures:

$$\mathbf{w}' = \mathbf{w} \odot (1 - \alpha \cdot \mathbf{p}), \tag{25}$$

where $\odot$ denotes element-wise multiplication.

*Final Weight Normalization.* Finally, the adjusted structure weights are normalized using the softmax function to obtain the final weight vector:

$$\mathbf{w} = \text{softmax}(\mathbf{w}'). \tag{26}$$

These steps ensure that the structure weights reflect both the importance of each structure and the redundancy between them, allowing for more effective fusion of the higher-order masks.

## E Loss Functions

In this section, we describe the loss functions used to enforce structural consistency in our model. Specifically, we define two types of losses: the Structural Verification Consistency Loss and the Integrated Mask Balance Loss.

*Structural Verification Consistency Loss.* The structural verification consistency loss evaluates the discrepancy between each higher-order structural mask and its corresponding verified version. It is defined as follows.

$$\mathcal{L}_t^{\text{verify}} = \sum\nolimits_{s\in\mathcal{S}} \omega_t^s \cdot \mathcal{D}_{\text{MSE}}(M_t^s, M_t^{s_{\text{verified}}}), \tag{27}$$

where $\omega_s^t \in \mathbb{R}^+$ is the importance weight of structure $s$ at time step $t$, $\mathcal{D}_{\text{MSE}}(\cdot,\cdot)$ is the mean squared error (MSE) distance function, defined as $\mathcal{D}_{\text{MSE}}(\mathbf{A},\mathbf{B}) = \frac{1}{n}\sum_{i=1}^{n}(a_i - b_i)^2$, where $n$ is the vector dimension, and $a_i$ and $b_i$ are the $i$-th elements of vectors $\mathbf{A}$ and $\mathbf{B}$, respectively.

*Integrated Mask Balance Loss.* The higher-order mask balance loss evaluates the consistency between the final higher-order causal mask and each verified mask, ensuring that the fused result properly integrates information from different structural views. It is defined as:

$$\mathcal{L}_t^{\text{balance}} = \sum\nolimits_{s\in\mathcal{S}} \mathcal{D}_{\text{MSE}}(M_t^{HC}, M_t^{s_{\text{verified}}}). \tag{28}$$

This loss encourages the higher-order fused mask to remain aligned with all structure-specific verified masks, promoting a balanced integration of multi-structural causal knowledge. By minimizing the deviation between $M_t^{HC}$ and each verified component, the model learns to generate unified yet structure-aware explanations.

## F Construction of Causal Representations

To obtain the causal representations used in our model, we apply Graph Convolutional Networks (GCNs) to the corresponding causal subgraphs at each time step. Specifically, four GCNs with learnable parameters $\Psi_{DC}$, $\Psi_{SC}$, $\Psi_{DHC}$, and $\Psi_{SHC}$ are employed to extract the following representations:

$$\begin{aligned} H_t^{DC} &= GCN(A_t^{DC}, X_t; \Psi_{DC}), \\ H_t^{SC} &= GCN(A_t^{SC}, X_t; \Psi_{SC}), \\ H_t^{DHC} &= GCN(A_t^{DHC}, X_t; \Psi_{DHC}), \\ H_t^{SHC} &= GCN(A_t^{SHC}, X_t; \Psi_{SHC}), \end{aligned} \tag{29}$$

where $A_t^*$ denotes the adjacency matrix of the corresponding causal subgraph, $X_t$ is the node feature matrix at time step $t$, and $\Psi_*$ represents the trainable parameters of each GCN module.

## G Experimental Settings

### G.1 Datasets

For the node classification task and the graph classification task, we use 4 synthetic datasets and 2 real-world datasets. The statistics of all datasets are shown in Table 2. To remain consistent with previous work [33], for the node classification task, we use a real-world dynamic graph dataset Elliptic2, and construct three synthetic datasets BA-Shapes, Tree-Cycles, and Tree-Grid [27] into dynamic graph datasets named DBA-Shapes, DTree-Cycles, and DTree-Grid, respectively. For the graph classification task, we use a real-world dataset MemeTracker [21], and convert the benchmark synthetic dataset BA-2motifs [11] into a dynamic graph dataset named DBA-2motifs. These datasets are selected to comprehensively evaluate our model's capability in handling both synthetic and real-world dynamic graphs. Synthetic datasets allow us to examine the model's ability to capture specific structures and causal patterns in a controlled environment, while real-world datasets validate its practical applicability under complex temporal and structural variations.

### G.2 Baselines

Since there are few existing works on explaining DyGNNs, in addition to comparing with DyGNNExplainer [33], we mainly compare our method with several powerful static GNN interpretability frameworks, including GNNExplainer [27], PGExplainer [11], Gem [8], and OrphicX [9]. For these static explanation methods, all nodes and edges in the graph are treated as occurring simultaneously.

### G.3 Experimental Setup

All experiments were conducted on a Windows-based workstation equipped with an NVIDIA A40 GPU, a 15-core Intel Xeon Platinum 8358P CPU, and sufficient memory. The model was trained for 200 epochs using the Adam optimizer with a fixed learning rate of 0.001. The dataset was split into training, validation, and test sets with a ratio of 7:1.5:1.5. During training, we adopted a composite loss function consisting of multiple components, including contrastive loss $\mathcal{L}_c$, sparsity loss $\mathcal{L}_s$, higher-order sparsity loss $\mathcal{L}_{hs}$, prediction loss $\mathcal{L}_p$, higher-order prediction loss $\mathcal{L}_{hp}$, dynamic consistency loss $\mathcal{L}_d$, higher-order dynamic loss $\mathcal{L}_{hd}$, and structural consistency loss $\mathcal{L}_{\text{high}}$. The weights for each component in the overall loss function were set as follows: $\xi_1 = 0.8$, $\xi_2 = 0.2$, $\xi_3 = 0.2$, $\xi_4 = 0.8$, $\xi_5 = 0.5$, $\xi_6 = 0.6$, $\xi_7 = 0.4$, and $\xi_8 = 0.4$. We performed discrete sampling for the hyperparameters $\xi_1$ to $\xi_8$ within the range of $[0.1, 1.0]$, and selected the final configuration based on the average performance on the validation set. Additionally, to ensure the stability and reliability of the results, each experiment was independently repeated 10 times with different random seeds, and the average performance across these runs is reported. In addition, the experiments were implemented using Python 3.9 and PyTorch 2.x, with PyTorch Geometric and scikit-learn as the primary libraries.

## H Case Study of CausalSKyHop

This case study illustrates the hierarchical explanation framework implemented in our approach, as depicted in Figure 9. The architecture operates through two complementary higher-order pathways: the motif VGAE branch extracts local semantic patterns through causal mask estimation and verification, while the hypergraph VGAE branch captures global multi-node dependencies via hypergraph modeling and higher-order dynamic masking. Both

**Table 2: Statistics of datasets for both node and graph classification.**

| Dataset | Node classification | | | | Graph classification | |
|---|---|---|---|---|---|---|
| | DBA-Shapes | DTree-Cycles | DTree-Grid | Elliptic | DBA-2motifs | MemeTracker |
| #nodes | 700 | 871 | 1,231 | 203,769 | 25,000 | 3.3 mil. |
| #edges | 4,110 | 1,950 | 3,410 | 234,355 | 51,392 | 27.6 mil. |
| #labels | 7 | 3 | 3 | 2 | 3 | 2 |

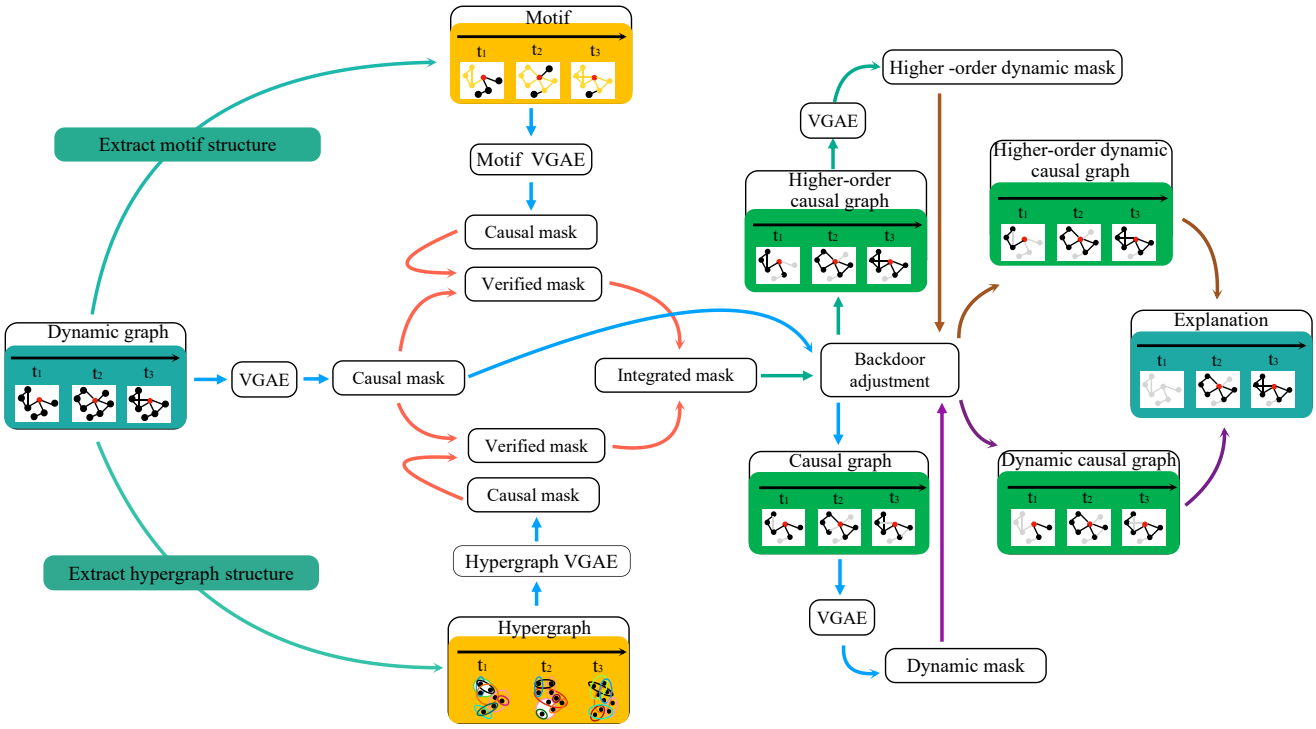



**Figure 9: The case study of CausalSKyHop. Using the red node as an example, the figure follows the pipeline "structure extraction → dual higher-order modeling branches → mask fusion and adjustment → dynamic causal graph and explanation outputs." From left to right, the panels correspond to the input, intermediate processing states, and the final results ($t_1 \rightarrow t_3$).**

pathways are enhanced with dedicated causal discovery mechanisms—including *causal mask*, *verified mask*, and *integrated mask*—to ensure structural and semantic coherence. The synergistic integration of motif- and hypergraph-based reasoning, supported by *backdoor adjustment* techniques, enables the construction of a unified *higher-order causal graph* that systematically disentangles genuine causal relationships from spurious correlations. This multi-granular masking strategy ensures that the final *explanation* is both semantically interpretable and temporally consistent, validating the importance of integrating complementary higher-order structures in dynamic graph interpretation.